Structures for establishing a database for marine monitoring

DOC SCIENCE INTERNAL SERIES 58

Marti J. Anderson

Published by
Department of Conservation
P.O. Box 10-420
Wellington, New Zealand

DOC Science Internal Series is a published record of scientific research carried out, or advice given, by Department of Conservation staff, or external contractors funded by DOC. It comprises progress reports and short communications that are generally peer-reviewed within DOC, but not always externally refereed. Fully refereed contract reports funded from the Conservation Services Levy are also included.

Individual contributions to the series are first released on the departmental intranet in pdf form. Hardcopy is printed, bound, and distributed at regular intervals. Titles are listed in the DOC Science Publishing catalogue on the departmental website http://www.doc.govt.nz and electronic copies of CSL papers can be downloaded from http://csl.doc.govt.nz

ISSN 1175-6519

ISBN 0-478-22285-8

This is a client report commissioned by Northern Regional Office and funded from the Unprogrammed Science Advice fund. It was prepared for publication by DOC Science Publishing, Science & Research Unit; editing and layout by Geoff Gregory. Publication was approved by the Manager, Science & Research Unit, Science Technology and Information Services, Department of Conservation, Wellington.

CONTENTS

Structures for establishing a database for marine monitoring

Marti J. Anderson

Department of Statistics, Tamaki Campus, University of Auckland, Private Bag 92-019, Auckland, New Zealand

ABSTRACT

A database structure for monitoring marine reserves is given, including actual data and meta-data to be kept on each individual study unit. The meta-data should contain all of the information necessary to understand the context and experimental design of the data, including variables measured, factors examined, spatial and temporal scales of sampling, and levels of replication. Protocols for monitoring should be established on the basis of knowledge of spatial and temporal variation at different scales for the variables of interest. This may include an analysis of precision as a function of sample size for each scale. As many aspects of the methods and designs as possible should be standardised so that statistical meta-analyses can be performed with broader, more rigorous inference. The essential elements for an appropriate experimental design for ongoing monitoring are: (a) 'before' and 'after' data collected from each reserve and at several reference locations per reserve; (b) sites within locations must be included and must be interspersed to avoid pseudo-replication; and (c) replication at several temporal and spatial scales will furthermore allow more powerful inferences concerning the extent of any impacts or effects of reserves. Statistical meta-analyses can be done on existing data to assess 'status and change' for certain variables in response to marine reserves. This can only be done, however, on a case-by-case basis (i.e. considering one variable or assemblage at a time) across the available studies. Extreme care must be taken to establish commensurability of sampling protocols in order to (a) put together internally consistent datasets for formal meta-analysis and (b) use an appropriate statistical model and method for the analysis and estimation of effects.

Keywords: Database design, monitoring, marine reserves, experimental design, sampling level, statistical meta-analysis.

Anderson, M.J. 2002. Structures for establishing a database for marine monitoring. *DOC Science Internal Series 58*. Department of Conservation, Wellington. 21 p.

1. Introduction

A database will be established by the Department of Conservation (DOC) for purposes of marine conservation management. In particular, it has been recommended that the Marine Survey and Monitoring Advisory Group (MSMAG): (i) obtain the raw data held within DOC and by external contractors; (ii) develop a storage system for the raw data; and (iii) decide how data will be used to determine condition and trend in marine reserves (McCrone 2001). The aim of this report is to provide advice as to a logical framework for such a database, from the perspective of a practising ecological scientist and statistician. To this end, the report responds to four specific questions required by DOC:

1. What is a database and what is a reasonable database structure for marine reserve biological monitoring?

2. Why should protocols for a monitoring database be standardised? How can these protocols be established for long-term monitoring for marine reserves?

3. What are the essential elements of experimental design needed to detect and assess 'status and change in marine reserves'?

4. Can existing raw data be used to assess 'status and change in marine reserves' and, if so, how?

The advice provided here is intended as only a starting point. Ultimately, the design, development and implementation of the database, issues of access and organisation, will come down to a database manager or software developer, who will have the skills to create the database, with specifications as required by DOC. Furthermore, no specific statistical analyses of any existing data will be pursued here. The report covers the important conceptual issues to be considered in the task of developing appropriate structures for a database, sampling design protocols for marine monitoring, and meta-analysis of existing data.

2. Database structure for marine reserve biological monitoring

A database is a structured array (or series of arrays) of information which is used to organise and store large quantities of information in such a way that specific information (i.e. subsets of the arrays) can be searched, found, retrieved, and/or used for analysis.

2.1 BASIC STRUCTURE

In the context of arrays of information for biological monitoring of marine reserves, the database should consist of individual *units*, with each unit corresponding to a particular study. Each unit needs to be numbered (or indexed in some other way), such as is commonly used for libraries of information. Each unit should consist of (at least) two parts (e.g. Fig. 1): (1) the actual array of data for a particular study (i.e. the 'data array') and (2) an array of information necessary to understand the content of the data array (i.e. the 'meta-data'). These two units should be identifiable with the same indexing unit number and should be inextricably linked in the database.

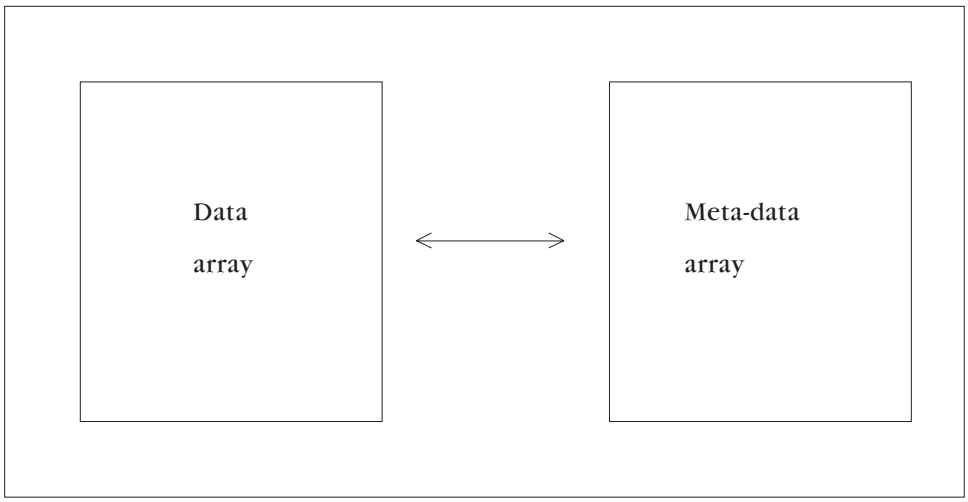

Figure 1. Every unit of the proposed database should have two components that are inextricably linked: the array of actual data and the array of meta-data that contains all of the information necessary to understand the content of the data array.

2.2 STRUCTURE OF META-DATA

A reasonable structure for the meta-data is:

1. A list of headings (which I shall refer to as 'fields') which are common across all data sets and may include (but not be limited to) the following:

   Unit No. The index or number for the particular unit in the database.

   Unit Name. An identifying name for the unit of study.

   Objectives. General and/or specific objectives of the study.

   Reference(s). Full citation of report(s) (and optionally, any published papers) for the study.

   Year(s) in which study was conducted.

   Location(s) or *Place(s)* where study was conducted.

   Habitat, e.g. subtidal rocky shore, estuarine intertidal, sandy beach, etc.

   Personnel who have done the work, along with their affiliation (other contact details may be included optionally).

2. Experimental design, which will have several sub-fields:

   Response variable(s) of interest: names of variables measured (e.g. Lobster abundance).

   Other variables measured: names of covariables or predictor variables (e.g. temperature or salinity).

   Temporal factors: name, categories, number of categories and frequency or scale.

   Spatial factors: name, categories, number of categories and scale.

   Other factors: name, categories, number of categories.

   Replication: name, size and number of replicates.

   Notes: any other additional relevant notes or information concerning methodology, context, relationships among factors, etc.

3. Keywords: Helpful words or phrases to identify the study unit when performing searches.

2.3 STRUCTURE OF THE DATA ARRAY

In general terms, the actual array of data will consist of a matrix, where columns give the names of variables and factors in the design and rows consist of values obtained for individual replicate observations for each variable with corresponding category names for each factor. One may swap columns and rows, but it is a good idea to keep this consistent across the entire database.

In addition, factor names, names of categories, names of variables, etc., as listed in the meta-data should correspond exactly to the corresponding names in the actual data array. For example, if the variable measured is called 'temperature' in the actual data array, it should be called 'temperature' (and not 'Temp' or 'degrees C') in the meta-data. This is necessary in order to maintain consistency and clarity in the link between the two arrays for each unit.

An example of a data array and its corresponding meta-data array is given in Fig. 2, using an excerpt of the data set provided for the Cape Rodney to Okakari Point survey for Rock Lobster conducted by Shane Kelly in 2000. In this particular case, there were no temporal factors, because the survey was done at only one time. In many cases, however, surveys are repeated at several times. For a temporal factor, 'scale' refers to the frequency of sampling. Thus, one may have a temporal factor named 'season' that has two categories 'Sep' and 'Mar,' which would correspond to a scale of 'biannual.' Also note that for these data, the 'Other variables measured' were categorical, rather than being quantitative. In many cases, the other variables measured will be quantitative variables like temperature, salinity, pH, or dissolved oxygen, etc. in which case it is important to also give the units of the measure (e.g. degrees C, ppt, %, etc.).

The difference between 'Other factors' and 'Other variables measured' is an important distinction. An experimental design is structured according to a factor, but not by reference to a measured variable. For example, if one measures lobster abundance in each of several transects at each of three depths, then depth is a structured 'other factor' in the design, with three levels. On the other hand, if one measures and records the depth at each transect (as well as

Actual data array

Status	Site	Transect	Lobsters	Depth	Censor
Reserve	One Spot	1	0	Deep	Tim
Reserve	One Spot	2	1	Deep	Tim
Reserve	One Spot	3	3	Deep	Shane
Reserve	One Spot	4	8	Deep	Shane
Reserve	One Spot	5	19	Deep	Shane
Reserve	Inner Table Top	1	54	Shallow	Shane
Reserve	Inner Table Top	2	20	Shallow	Shane
Reserve	Inner Table Top	3	2	Shallow	Tim
Reserve	Inner Table Top	4	3	Shallow	Tim
Reserve	Inner Table Top	5	1	Shallow	Tim
Reserve	Knot Rock	1	1	Shallow	Shane
Reserve	Knot Rock	2	0	Shallow	Shane
Reserve	Knot Rock	3	1	Shallow	Shane
Reserve	Knot Rock	4	4	Shallow	Tim
Reserve	Knot Rock	5	9	Shallow	Tim
Reserve	Outer Table Top	1	0	Deep	Shane
Reserve	Outer Table Top	2	4	Deep	Shane
Reserve	Outer Table Top	3	4	Deep	Tim
Reserve	Outer Table Top	4	8	Deep	Tim
Reserve	Outer Table Top	5	0	Deep	Tim
Reserve	Outer Martins	1	1	Deep	Shane
Reserve	Outer Martins	2	15	Deep	Shane
Reserve	Outer Martins	3	0	Deep	Shane
Reserve	Outer Martins	4	6	Deep	Tim
Reserve	Outer Martins	5	0	Deep	Tim
Reserve	Inner Martins	1	4	Shallow	Shane
Reserve	Inner Martins	2	40	Shallow	Shane
Reserve	Inner Martins	3	16	Shallow	Shane
Reserve	Inner Martins	4	9	Shallow	Tim
Reserve	Inner Martins	5	5	Shallow	Tim
Non-Reser	Outer Slater Soutl	1	0	Deep	Shane
Non-Reser	Outer Slater Soutl	2	0	Deep	Shane
Non-Reser	Outer Slater Soutl	3	0	Deep	Shane
Non-Reser	Outer Slater Soutl	4	0	Deep	Tim
Non-Reser	Outer Slater Soutl	5	0	Deep	Tim
Non-Reser	Inner Slater South	1	0	Shallow	Shane
Non-Reser	Inner Slater South	2	0	Shallow	Shane
Non-Reser	Inner Slater South	3	0	Shallow	Tim
Non-Reser	Inner Slater South	4	0	Shallow	Tim
Non-Reser	Inner Slater South	5	0	Shallow	Tim
Non-Reser	Outer Slater North	1	9	Deep	Shane
Non-Reser	Outer Slater North	2	10	Deep	Shane
Non-Reser	Outer Slater North	3	2	Deep	Shane
Non-Reser	Outer Slater North	4	7	Deep	Tim
Non-Reser	Outer Slater North	5	2	Deep	Tim
Non-Reser	Inner Slater North	1	0	Shallow	Shane
Non-Reser	Inner Slater North	2	0	Shallow	Shane

Meta-data array

Unit No.	1				
Unit Name	Cape Rodney to Okakari Point Rock Lobster Survey 2000				
Objectives	Compare abundances of rock lobsters inside vs. outside the marine reserve				
Reference(s)	Kelly, S. 2000. Cape Rodney to Okakari Point Marine Reserve lobster monitoring progr				
Year(s)	2000				
Location(s)	Cape Rodney to Okakari Point				
Habitat	subtidal rocky reefs				
Personnel	Shane Kelly, Tim Langlois				
Affiliation	Coastal & Aquatic Systems Limited				
Experimental design:	**name**	**categories**	**no. categories**	**scale**	**units**
Response variable(s):	Lobsters				counts
Other variables measured:	Depth	Deep, Shallow	2		
	Censor	Tim, Shane	2		
Temporal factors:					
Spatial factors:	Sites	One Spot, Inner Table Top,	12	10's to 100's of m	
Other factors:	Status	Reserve, Non-Reserve	2		
Replication:	Transects		5	50m x 10m	
Notes:	Sites were a random factor, with 6 sites nested in each of the Status categories. Trans				
Keywords:	MR monitoring, Jasus edwardsii, CROP, population dynamics				

Figure 2. An example of an actual data array and its accompanying meta-data, using an excerpt of the data set provided for the Cape Rodney to Okakari Point survey for Rock Lobster conducted by Shane Kelly in 2000. Note that, in the above printout, the cells for some fields in the Excel spreadsheet for the meta-data do not visually display all of the information they actually contain, but this is simply due to formatting.

measuring the variable of interest, which is lobster abundance), then depth is simply an 'other variable measured.'

2.4 SEARCH AND RETRIEVAL: CONSISTENCY IN ADDING TO AND MAINTAINING THE DATABASE

A person should be able to search any specified field of the meta-data and obtain a list of the studies matching a particular specification for that field. For example, I should be able to search 'Response variable' = 'Lobsters' and obtain a list of all study units that measured lobsters. Obviously, as anyone who has used the web to search for items from a library or database will be aware, search engines may be constructed to handle more complex requests, such as particular combinations of terms in more than one field at a time.

It is important that, when creating, adding to and maintaining the database, the *specific terms* that are used (for names of variables, scales of factors, habitat, etc.) remain consistent between actual data arrays and meta-data arrays and across the entire database so that search and retrieval with regard to particular

terms in specific fields can be done easily and efficiently. For example, it should be possible to search for 'Transect' in the field for 'Replication' and '50 m × 10 m' in the field for 'scale' to obtain a list of all study units that used transects of a particular size.

For search purposes, it will also be useful to have a particular field in the meta-data simply called 'Keywords' that will contain useful terms for purposes of the kinds of searches that may be done routinely. For example, DOC may find keywords or phrases such as 'MR monitoring,' 'MPA,' 'Human Impact Study,' or 'Rehabilitation' useful for search and identification purposes. As for the other fields, the more consistent these kinds of keyword terms can be throughout the database, the better.

2.5 MERITS OF PROPOSED DATABASE STRUCTURE

The above structure is only intended as a suggested template, which may be modified to suit DOC's management needs. The merits of the proposed database structure are:

- It has a simple, consistent structure, flexible enough to handle data from any ecological study.
- It will allow searches to be done on the basis of virtually any aspect of logical interest for marine monitoring.
- The fundamental units can be extended to have a greater number of arrays (e.g. there may be links created within the unit to the arrays of the original data file(s) provided by contractors, etc.).
- Structured in this way, fundamental aspects of the experimental design are clear so that scientists and statisticians may easily extract relevant data from individual studies for meta-analysis.
- The structure of the database is easy to modify by the addition or modification of existing fields.

It is important to emphasise that the *analysis* of data should be kept *separate* from the *database* itself. As nice as it would be for managers to have the ability to 'push a button' to obtain answers to relevant questions of interest, combining information from existing data for analysis is fraught with difficulties and should be done on a case-by-case basis by statisticians in response to specific questions and hypotheses. In short, it is simply not possible to have a database that also asks and answers important questions of the data as well. Statisticians and scientists should be employed for this task.

However, once consistent protocols have been established for specific sampling designs in ongoing monitoring programmes (e.g. existing sampling designs are done consistently and routinely for certain organisms in certain areas, with new data each year), it may be possible to develop computer programmes with analytical tools that may be used routinely for analysis. This is a topic for future discussions.

3. Protocols for a monitoring database

Protocols (i.e. experimental designs and methods for sampling) should be standardised as much as possible. This means studies involving the measurement and estimation of populations, habitats or assemblages should be done:

- With the same or similar size and shape of replicate units (e.g. cores, transects, surveys or quadrats).
- With the same or similar level of replication (i.e. sample size, *n*).
- With the same or similar number and scale of spatial and temporal factors (e.g. data collected annually at sites separated by hundreds of metres, etc.).
- With the variables measured in the same or similar ways (e.g. counts by visual census done over the same sized area, by the same personnel, if possible).
- Of the same species variables, if more than one species is measured.

In short, studies that attempt to measure the same response variable (e.g. snapper abundance) or set of response variables in an assemblage (e.g. estuarine intertidal soft-sediment communities) at different places and repeatedly through time should be as similar as possible in all aspects of their experimental design and sampling protocol.

Why is this important? There can be no doubt that the best way to obtain rigorous statements about effects on a large scale (i.e. across large geographic areas) and/or in the long run (i.e. over long periods of monitoring through time) is to set up large temporal and spatial scales as independent factors to be modelled in a statistical analysis. This clearly cannot be done in any sensible way if the individual studies are not commensurable with regard to their method and design. Commensurability is paramount in order for any 'big picture' questions to be assessed.

3.1 PRECISION: USE OF PILOT STUDIES FOR DEVELOPING PROTOCOLS

In terms of developing protocols to be maintained for individual studies, some considerable effort will be required to establish appropriate sample sises and levels of replication at the right spatial and temporal scales for the particular variable(s) of interest. This will require some care in the early stages of monitoring through the use of pilot studies or analyses of existing studies. One can use a pilot study to measure and estimate the variability at various different spatial and temporal scales, which will indicate the amount of replication needed and at which scales of measurement (e.g. Andrew & Mapstone 1987; Underwood 1997).

Precision can be defined as the degree of concordance among a number of measurements or estimates for the same population (Andrew & Mapstone 1987).

A good measure of precision is given by:

$$p = SE/\bar{x} = (s/\sqrt{n})/\bar{x}$$

where $\bar{x}$ is the mean of a sample with n replicates, s is the standard deviation and $SE = s/\sqrt{n}$ is the sample standard error. Note that as the value of p decreases, precision increases. Although variance is independent of sample size (i.e. it is a characteristic of the population), precision reflects the sampling programme and increases with sample size. Within one of the sampling areas to be used for the investigation, for example, one may start by obtaining a sample with a relatively large sample size, n. Having obtained these data, one can then examine random subsets of the data set (i.e. where the sample sises are n – 1, n – 2, …, 2). For each of these one can estimate the precision and draw a plot like the one shown in Fig. 3, in which the original sample size was n = 10 and estimates of precision were obtained for random subsets of the observations where n = 9, 8, …, 2. For this species (a polychaete worm, *Prionospio pinnata*, in cores of soft sediment from the Okura estuary), precision starts to level off around n = 5, which also achieves a precision of less than 0.5 (Fig. 3).

As n increases, we expect a gradual decrease to a levelling off in the value of p. Where p begins to level off indicates a reasonable sample size to use for the larger study, as using larger values for n would not result in particularly great increases in precision.

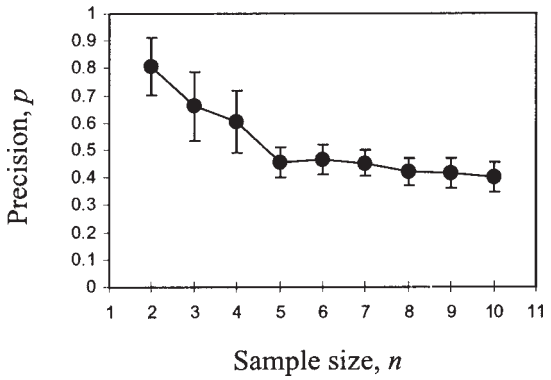

Figure 3. Example of a plot of precision versus sample size.

Another approach is to set the desired precision *a priori*. For example, a precision of 0.5 is generally considered to be reasonable for most ecological investigations. By setting a desired precision *a priori* (i.e. p = 0.5 or p = 0.1), the researcher can choose an appropriate sample size, given preliminary estimates of s and $\bar{x}$ from a pilot study. For a desired precision (p), an appropriate sample size (n) is given by $n = [s/(p\bar{x})]^2$. One can investigate appropriate replication for any spatial or temporal scale in the experimental design in a similar fashion. Analogous measures can also be used for establishing appropriate sample sizes for multivariate response data (Anderson et al. 2001).

4. Essential elements of experimental design

What are the essential elements of experimental design needed to detect and assess 'status and change in marine reserves'? The answer to this question will depend on (a) which variable(s) is (are) being targeted and (b) the degree of natural variability in that variable of interest at different spatial and temporal scales. That is, first of all we need to define more specifically *what it is* about the marine reserve we consider important to understand by doing the monitoring. Second, an appropriate sampling design can be made by reference to knowledge of the natural levels of variability for that targeted species or set of variables.

4.1 WHAT TO MONITOR

In some cases, the variable of interest will be a particular species, such as the rock lobster or paua. This might be because the public have a vested interest (for commercial, ethical, social or spiritual reasons) in that species. In other cases, we may be concerned with the status or potential change in a whole assemblage of species simultaneously, such as an intertidal rocky shore community, which is a set of interacting multivariate response variables.

In many cases, changes due to the establishment of marine reserves will have both direct and indirect effects on many species of organisms. For example, it has been demonstrated that establishment of marine reserves has led to an increase in large predatory fishes, including snapper, inside reserves. Predation by these fish has caused decreases in populations of sea urchins (kina), which, in turn, has led to increases in the density and extent of kelp forests inside marine reserves (Babcock et al. 1999). Thus, the particular variables (or sets of variables) to be monitored must be articulated carefully and must be understood by reference to the 'bigger picture,' i.e. the interactive system they exist within.

This is why, in addition to monitoring specific 'target' species of interest, it is also important to monitor whole sets of species (whole systems) simultaneously, as far as possible. Indeed, multivariate methods (e.g. Clarke 1993; Anderson 2001), involving the simultaneous analysis of many interacting response variables (species), are pivotal in this context, as they will have greater power than single indicator species, and can also detect different kinds of environmental impacts (Underwood & Peterson 1988; Clarke 1993).

4.2 EXPERIMENTAL DESIGN FOR MONITORING

In the scientific literature, there have been many developments in the area of experimental designs to assess environmental impact. The simplest design involves what is called 'Intervention analysis,' where only the impact location is monitored through time and the assessment consists of comparing the time series for the response variable(s) *before* the purported impact with that occurring *after* (Box & Tiao 1965, 1975). Next in line is the traditional 'BACI' analysis, where the changes through time 'Before' and 'After' for a 'Control' location are compared with those for an 'Impact' location (Green 1979; Stewart-Oaten et al. 1986). A further modification to the BACI design included the use of multiple control locations and asymmetrical analyses (Underwood 1991, 1992, 1994). I shall refer to the latter as 'IVRS' designs, for 'impact versus reference sites', after Stewart-Oaten & Bence (2001).

I had recently observed that the IVRS designs tended to have very little power to detect important changes in natural systems. I had determined several important reasons for this: (i) the number of reference locations dictates the power of the test for impact, and these are usually limited (e.g. two reference sites would result in only one error degree of freedom for the statistical test); (ii) high natural variability among random reference locations is common and further decreases power; (iii) the treatment of 'locations' as a random factor causes loss of power for the design and, furthermore, is logically unrealistic, as locations are actually usually chosen quite carefully and explicitly; (iv) the treatment of several times of sampling as 'random' is unrealistic, as time has a single direction and there are often correlations through time in biological variables at various scales.

Most recently, Stewart-Oaten & Bence (2001) and Hewitt et al. (2001) have challenged the IVRS models and suggested that better and more powerful models can be used. Some of their suggestions include: (i) not treating the locations as random; (ii) using several reference locations that are as highly correlated with the purported 'impact' location as possible; (iii) treating the reference locations as covariates in the analysis; and (iv) using the variability through time as the random error, rather than the variability among locations used by the IVRS approach.

It is worth pointing out that, in the context of marine reserves, interest may actually lie in seeing how the ecosystem changes as a consequence of the establishment of a reserve. Thus, rather than detection of 'impact', managers may rather be interested in how the reserve is doing in terms of changes towards a 'less polluted' or 'natural' state*. In any case, the same basic principles essentially apply in designing a study to detect changes to become more 'natural' as opposed to changes to become more 'impacted.'

---

* Of course, defining what is meant by 'natural' is always important in this context and may need to be defined by reference to some long-term existing reserves, etc.

4.3 RECOMMENDATIONS ON ESSENTIAL ELEMENTS

Taking into consideration all of the available literature to date on the subject, I recommend that the following essential elements of experimental design be included in a monitoring programme:

- 'Before' data should be collected and replicated through time. That is, data should be collected (and at several times) before the potential impact occurs or (as the case may be) before the establishment of a new marine reserve.
- 'After' data should be collected and replicated through time. This is imperative for ongoing monitoring. Even if 'Before' data are not available, 'After' data should be maintained, as there are some analytical mechanisms available for dealing with this problem (Glasby 1997).
- Spatial reference sites should also be monitored (i.e. at the same sampling times). Reference sites should be as similar as possible in all relevant ways to the potentially impacted (or reserve) site (e.g. similar wave exposure, salinity regime, etc.). Although, at the barest minimum, one can analyse data with only one spatial reference site, the number of spatial reference sites should be increased to as many as is practically possible. This will greatly increase the power to detect any impacts (or other changes).
- Reference sites should be spatially interspersed with target (or potentially changing or impacted) sites. This is extremely important to avoid pseudoreplication (Hurlbert 1984). If reference sites are systematically spatially segregated from target sites, there will be no sure way of attributing any potential differences to causes other than spatial variance.

In addition, the design and the potential information content of outcomes from the monitoring programme may be greatly enhanced by considering the following:

- Replication of studies at the level of whole reserves (i.e. several target sites or locations) will allow much wider statistical inferences concerning any effects (e.g. Beck 1997). That is, one may be able to consider effects of 'marine reserves' rather than being restricted to only considering effects of a particular reserve at a particular time and place.
- Including spatial replication *at several scales* can be used to give an indication of the *spatial extent* of any impact (or other changes). That is, an impact might only be occurring locally and may not affect areas on a larger scale. With only one spatial scale of sampling, it is not possible to estimate the spatial extent of any impacts.
- Including temporal replication *at several scales* can be used to distinguish 'pulse' (short-term) from 'press' (long-term) effects, i.e. to measure the temporal extent of any impacts.

4.4 CONTROL CHARTS AS A TOOL FOR MONITORING

It may be useful to distinguish the activity of 'monitoring' from that of 'environmental impact assessment.' More often than not, unless the impact is patently obvious (such as an oil spill), it is usually not known where or when it may have occurred or may still be occurring. Indeed, one might argue that the role of monitoring (as the name suggests) should be to provide a 'signal' or 'alarm bell' to the presence of an impact, if, where, and when it does occur. That is, we should hope that a reasonable monitoring program would provide us with a way of assessing, at any particular time, whether a measurement we observe is unusual, given what we would expect from our observations of the naturally variable system until that time.

Sequential statistical methods, such as cumulative sums (CUSUMs), sequential probability ratio tests (SPRTs) and control charts, as developed for industrial applications, offer some promise in this regard (e.g. Shewhart 1931; Wald 1947; Wetherill 1975; Montgomery 1996). These methods were originally developed to provide a way of identifying when a system (e.g. in a factory or other industrial context) was going 'out of control', so as to trigger an alarm to stop the system and employ appropriate remedial measures. Control charts involve plotting through time some measure of a random process by reference to its expected value (the mean) and specified upper and lower bounds for the measure. If a control chart measure exceeds the allowable bounds (usually specified as three standard deviations from the mean), this indicates a shift in the target value; that is, that something in the system has gone awry (e.g. Montgomery 1996).

Green (1979) mentioned sequential methods in his seminal book for environmental biologists, but there are relatively few examples of their use in monitoring biological populations and communities (but see Schipper et al. 1997 and Pettersson 1998). There may be several reasons for this, but one is that biological variables do not behave as normal random variables, and so special non-parametric statistical methods are needed.

Working with the Australian Institute of Marine Science, which has the extremely difficult task of monitoring the entire Great Barrier Reef over long periods of time, I have recently developed new non-parametric statistical methods for producing multivariate control charts. This would be a useful approach to consider in the future for marine monitoring here in New Zealand as well, although the work (and its accompanying computer program) has not yet been published for general use (M.J. Anderson and A.A. Thompson, unpubl. data).

Two important points should be considered if some kind of control charts are to be used for monitoring: (i) the monitoring design (places monitored, level of replication, etc.) for any particular species or sets of variables must remain consistent through time and across reserves and (ii) a reasonably long 'run-in' time is needed before the appropriate target value (mean) and its variance can be estimated.

5. Use of existing raw data

5.1 BACKGROUND: META-ANALYSIS

The assessment of the weight of existing evidence for or against a particular effect or hypothesis (and measurement of the size of said effect) is generally done through the use of what is called a 'meta-analysis'*. This term was coined by Glass (1976) and refers to the synthesis of evidence from several investigations or experiments. The traditional approach to meta-analysis was done simply by experts reading, considering and weighing existing information, which they then generally summarised as a review of the available literature on the topic of interest. Another early subjective approach, called 'evidence tables' by Eddy (1992), was to list for each study, in tabular form, the design, sample size, variables measured and results obtained. Some subjective statements concerning the content of the evidence table could then be made.

However, such subjective approaches are no longer readily accepted in scientific circles. The development of rigorous statistical methods for meta-analysis (combining information from several studies) has become a field unto itself. One of the earliest statistical methods for meta-analysis was provided by Fisher (1932), who gave a formula for combining and analysing *P*-values from several independent studies. The advantage of this approach is that the individual studies do not need to have the same experimental design. However, an important drawback is that the studies are not weighted in any way for their degree of uncertainty or sample size, nor does this approach provide any estimates of the sizes of effects (Hasselblad 1994).

More recent techniques include the use of random effects models. Here, the variation is partitioned into two sources or parts: (i) variation due to the random variability within the particular study itself and (ii) variation due to the random variability from study to study (Hedges & Olkin 1985; DerSimonian & Laird 1986). This general approach is particularly useful and robust and can be used with great efficacy, most particularly if studies of the same organism(s) are done at several times or places using the same (or similar) experimental designs.

More complicated methods of meta-analysis exist and can be used to advantage, such as Bayesian approaches (e.g. Eddie 1989; Tweedie et al. 1996) and non-parametric techniques (Duval & Tweedie 2000). These methods are particularly useful for taking into account any biases that may be known to be present for existing studies†, or to allow different aspects of the variation between studies to be modelled in better ways than the random effects model.

---

* It is important not to confuse the term 'meta-analysis' with the term 'meta-data,' which was discussed above.

† For example, a well-known problem is the so-called 'publication bias,' which refers to the fact that only studies that show significant effects tend to get published. See Duval & Tweedie (2000) for more details on recent developments.

Meta-analysis becomes much more 'do-able,' and with much better and more interpretable outcomes, when the studies upon which it is based have the same structure, experimental design and objectives. In contrast, a meta-analysis, of necessity, becomes much more complicated, less interpretable and more problematic the more incommensurable the individual studies are.

5.2 AN EXAMPLE: LOBSTER ABUNDANCE

Let us consider an example of a reasonable meta-analysis that might be done from some of the data provided. First of all, the primary targeted species for monitoring appear to be commercially important species, such as lobster, paua, or snapper. For these individual species, appropriate univariate models may be built, using meta-analysis, to estimate the effects of the reserves to date on each of these species. In addition, multivariate models may be built for multivariate data, such as the fish data from Tuhua, where abundance was recorded for a whole set of fish species as a multivariate response unit.

How can we go about building these models for an appropriate analysis? The first thing to do is to decide what is commensurable about the experimental designs and to determine which studies (or parts of studies) can be used together in the meta-analysis. A good way to do this is to list the information available about the studies in a table. For example, we can consider the data available on abundances of the lobster, *Jasus edwardsii*. Fortunately, it appears that similar methods were used in most of the surveys that have been done to date. For example, it would appear that all of the studies (except for Te Angiangi in 1995) have sampled several sites, inside and outside reserves, using transects measuring 50 m × 10 m as their units of replication. Table 1 lists the information available about key features of the design we will need to consider.

The following experimental design could be analysed for these data:

Factor 1: *Study* (specific to a certain time and place), 10 levels, random;

Factor 2: *Reserve Status*, two levels (inside v. outside), fixed, orthogonal to Factor 1;

Factor 3: *Sites*, nested in Factors 1 and 2, no. of levels = variable, random.

The no. of transects, *n,* is variable, depending on the *Study* × *Reserve Status* combination.

This is clearly an unbalanced design, both at the level of replication and at the level of individual sites, but it can be analysed as such. The question of interest would be whether there was a significant *Study* × *Reserve Status* interaction (indicating that the effect of the reserve depends on the particular study one chooses to look at). Also of interest would be to estimate the size of the main effect '*Reserve Status*,' given the random variability across all studies.

This sounds quite straightforward, but there are a number of important questions that will need to be answered carefully before we proceed with analysing or modelling the data in this way:

1. What is the scale of a '*Site*' for each study (i.e. how far apart are the sites)? Is this scale the same across the studies? If not, then the data may not be combined with '*Study*' expressed in a formal way as a factor.

TABLE 1. SUMMARY OF SOME DETAILS OF THE EXPERIMENTAL DESIGNS USED TO MEASURE EFFECTS OF MARINE RESERVES ON LOBSTER ABUNDANCE.

LOCATION	YEAR	NO. OF SITES INSIDE/OUTSIDE	NO. OF TRANSECTS PER SITE	DEPTH
The Gut	2000	3/3	5	7–17 m
Tuhua	1991	6/6	3 in each of 2 depths	Shallow/deep
	1993	6/6	3 in each of 2 depths	Shallow/deep
	1995	6/6	5 in each of 2 depths	Shallow/deep
	1998	6/6	3 in each of 2 depths	Shallow/deep
Te Angiangi	1996	20/20	?	?
	1998	11/13	?	?
	1999	16/12	?	?
	2000	6/6	?	?
CROP	2000	6/6	5	?

2. What was the spatial array of the '*Sites*'? For each study, were sites inside the reserves interspersed with the sites outside the reserves? If not, then pseudoreplication (*sensu* Hurlbert 1984) may be an issue.
3. Were the studies done at similar depths? What did 'shallow' and 'deep' for the Tuhua studies correspond to in real depth? Can information from the two depths be combined or not (i.e. is 'depth' a significant factor)? Was depth recorded for each transect (i.e. can depth be treated as a covariable)?
4. Were the studies standardised in some way for the available habitat for lobsters? A transect that does not have caves, holes or overhangs would not be expected to contain lobsters. Were the studies stratified for habitat or was suitable habitat measured in any way? This is a biological issue that has bearing on the statistical treatment of the data.
5. What was the spatial scale for transects (i.e. how far apart were transects within sites)? Is this consistent across the studies?

If it turns out that various aspects of the studies are found to be incommensurable, then a more indirect method will need to be used, rather than treating '*Study*' as a formal random factor in the analysis. This will probably involve estimating the size of the effect of the marine reserve (as well as its variance or error) for each study and combining these results to estimate a posterior distribution for the effect of marine reserves on Lobster abundance.

Whether we analyse '*Study*' as a random factor or do a meta-analysis of the effect of marine reserves, treating individual studies separately, there are still a few loose ends. There are, in fact, many ways that the individual studies differ, including the particular people doing the surveys, the year in which the surveys were done and the biogeographic and hydrodynamic differences among the locations. Thus, all of these things are confounded with the factor '*Study*' and there is no way of attributing any differences among the studies to any single one or more of these potential contributing causes. Also, in the above random effects model we have not formally incorporated any particular existing knowledge of the 'quality' of the studies into the analysis; rather, we have treated all of the studies equally. Furthermore, in the above model we have not taken into account any differences among the marine reserves that might be

very important to model, such as the age of the reserve (time since establishment). We also have not looked at modelling change, only at modelling the effect (if any) of marine reserves on lobster abundance across all times and places for which we have information. It is clear that complexities abound for setting up appropriate models for meta-analysis.

A formal way to incorporate our confidence in particular studies (i.e. if the survey methodology was better, the replication was greater, or the variability was smaller, etc.) would be to incorporate this knowledge formally into the prior in a Bayesian analysis. The details of this are too complex to cover here, but are certainly worth a mention. Bayesian analysis can be used to 'update' our existing beliefs about a parameter (such as the effect of a marine reserve on lobster abundance), using new data. One could implement this by doing a 'chain' of sequential Bayesian analyses on the existing data, rather than a single analysis. Further details are beyond the scope of this report. Suffice it to say here that a formal Bayesian approach, possibly in addition to the random effects models, would be a useful way to perform meta-analyses on the existing data for marine reserves.

Thus, in answer to the question of whether or not existing data can be used to assess the status and change of marine reserves, the answer is 'yes,' but it is clear that this will involve some complex analyses, specific to particular response variables and requiring careful consideration of the experimental designs of the studies. It will also not be without some problems, in terms of the statistical inferences to be drawn, until the studies are following a standard and repeatable protocol of survey design through time.

6. References

Anderson, M.J. 2001: A new method for non-parametric multivariate analysis of variance. *Austral Ecology 26*: 32–46.

Anderson, M.J.; Saunders, J.E.; Creese, R.G. 2001: Ecological monitoring of the Okura Estuary. Report I: Results of a pilot study. Unpubl. report, Auckland Uniservices Ltd, University of Auckland.

Andrew, N.L.; Mapstone, B.D. 1987: Sampling and the description of spatial pattern in marine ecology. *Oceanography and Marine Biology Annual Review 25*: 39–90.

Babcock, R.C.; Kelly, S.; Shears, N.T.; Walker, J.W.; Willis, T.J. 1999: Changes in community structure in temperate marine reserves. *Marine Ecology Progress Series 189*: 125–134.

Beck, M.W. 1997: Inference and generality in ecology: current problems and an experimental solution. *Oikos 78*: 265–73.

Box, G.E.P.; Tiao, G.C. 1965: A change in the level of a non-stationary time series. *Biometrika 52*: 181–192.

Box, G.E.P.; Tiao, G.C. 1975: Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association 70*: 70–79.

Clarke, K.R. 1993: Non-parametric analysis of multivariate changes in community structure. *Australian Journal of Ecology* 18: 117–143.

DerSimonian, R.; Laird, N. 1986: Meta-analysis in clinical trials. *Controlled Clinical Trials 7*: 177–188.

Duval, S.; Tweedie, R. 2000: A non-parametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95: 89–98.

Eddy, D.M. 1989: The confidence-profile method: a bayesian method for assessing health technologies. *Operations Research 37*: 210–228.

Eddy, D.M. 1992: Manual for evaluating health practices and designing practice policies. American College of Physicians, Philadelphia.

Fisher, R.A. 1932: Statistical methods for research workers, 4th edition. Oliver & Boyd, London.

Glasby, T.M. 1997: Analysing data from post-impact studies using asymmetrical analyses of variance: a case study of epibiota on marinas. *Australian Journal of Ecology 22*: 448–459.

Glass, G.V. 1976: Primary, secondary and meta-analysis of research. *Educational Researcher 5*: 3–8.

Green, R.H. 1979: Sampling design and statistical methods for environmental biologists. John Wiley & Sons, New York.

Hasselblad, V. 1994: Meta-analysis in environmental statistics. Pp. 691–716 *in*: Patil, G.P.; Rao, C.R. (eds) Handbook of statistics, Vol. 12: Environmental statistics. Elsevier Science, North Holland, Amsterdam.

Hedges, L.V.; Olkin, I. 1985: Statistical methods for meta-analysis. Academic Press, Boston.

Hewitt, J.E.; Thrush, S.E.; Cummings, V.J. 2001: Assessing environmental impacts: effects of spatial and temporal variability at likely impact scales. *Ecological Applications 11*: 1502–1516.

Hurlbert, S.H. 1984: Pseudoreplication and the design of ecological field experiments. *Ecological Monographs 54*: 187–211.

McCrone, A. 2001: National overview of biological monitoring in New Zealand's Marine Protected Areas. Northern Regional Office, Department of Conservation, Hamilton.

Montgomery, D.C. 1996: Introduction to statistical quality control, 3rd edn. John Wiley & Sons, New York.

Pettersson, M. 1998: Monitoring a freshwater fish population: statistical surveillance of biodiversity. *Environmetrics 9*: 139–150.

Schipper, M.; Den Hartog, J.; Meelis, E. 1997: Sequential analysis of environmental monitoring data: optimal SPRTs. *Environmetrics 8*: 29–41.

Shewhart, W.A. 1931: Economic control of quality of manufactured product. Van Nostrand, New York.

Stewart-Oaten, A.; Bence, J.R. 2001: Temporal and spatial variation in environmental impact assessment. *Ecological Monographs* 7: 305–339.

Stewart-Oaten, A.; Murdoch, W.W.; Parker, K.R. 1986: Environmental impact assessment: 'Pseudoreplication' in time? *Ecology 67*: 929–940.

Tweedie, R.L.; Scott, D.J.; Biggerstaff, B.J.; Mengerson, K.L. 1996: Bayesian meta-analysis, with application to studies of ETS and lung cancer. *Lung Cancer 14 Suppl. 1*: S171–S194.

Underwood, A.J. 1991: Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research 42*: 569–587.

Underwood, A.J. 1992: Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology 161*: 145–178.

Underwood, A.J. 1994: On Beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecological Applications 4*: 3–15.

Underwood, A.J. 1997: Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press, Cambridge, UK.

Underwood, A.J.; Peterson, C.H. 1988: Towards an ecological framework for investigating pollution. *Marine Ecology Progress Series 46*: 227–234.

Wald, A. 1947: Sequential analysis. John Wiley & Sons, New York.

Wetherill, G.B. 1975: Sequential methods in statistics. Chapman & Hall, London.